**COVER LETTER**

**Submission ID**: 128

**Title**: Interactive Exploration of Ligand Transportation through Protein Tunnels

**Date**: July 17th, 2016

We would like to thank the anonymous reviewers for their helpful comments. We attempted to address all remarks in the revised version of the paper. In this letter, we give detailed responses to all reviewer comments and describe the corresponding changes in our manuscript.

**Reviewer 3 - primary**

1. *There was one significant scientific concern raised by several reviewers, and ask that you do pay attention to that: specifically your choice of trajectory simplifications raises questions regarding its properties and suitability.  There is previous work on trajectory simplification, some quite well studied.  Two of the reviewers pointed out specific works that might be appropriate to consider. It would be helpful to your readers to at least put your algorithm in the context of previous (molecular) trajectory simplification work, or even to provide a choice of simplification algorithms, there are interesting differential benefits between them.*

TODO přidat komentář až poté, co se rozhodne, jestli se do článku vejde Douglas-Peucker nebo ne.

1. *From a vis perspective it could use some more work evaluating the correctness of what has been done, and whether there might be alternative solutions that are superior, but, from a biological perspective it's clear that the approach used in the manuscript "works", and having something that works out there for people who are trying to understand molecular ligand/docking trajectories, would be a real plus.*

The correctness of our solution was evaluated namely from the biochemical point of view. We aimed to explore if the tool is beneficial for the biochemists and if it gives them the valuable insight to the data. During the design phase we explored more possible visual representations and their applicability to our problem. Finally we concluded that the selected set of visualizations is the most appropriate and understandable by the domain experts.

1. *One area where I wish the manuscript contained more detail and comparison, is in the simplified trajectory material.  It appears that the simplification mechanism "works", but it's not obvious how well it works, or how robust it is to small variations in the input.  For example, it would be nice to know that the simplification produced from one set of MD runs for a ligand into a pocket, is similar to the simplification produced from a different set of MD runs for the same ligand and pocket.  It would also be nice to have more evidence that the simplification reliably preserves "interesting features" in the trajectory.  Admittedly, there are some challenges to performing these studies, but, whether in this paper or a follow-up, it would be good to give the reader some more information in this area.*

To address this issue, we added a case showing the stability of the trajectory simplification. We used a simulation of molecular dynamics and its artificially modified counterpart. In order to obtain the second simulation we have shifted every ligand position by a random vector in interval between (-1,-1,-1) (1,1,1).

TODO – trochu to rozšířit

**Reviewer 1**

1. *There are some minor questions on the design choices (see below). My main question would be about the feedback. Apparently, the biochemists have used the tool. By themselves? What was the major gain? New insight? Or just more efficient? By how much did the efficiency increase?*

As we aimed to design the tool to fit the biochemical needs, the cooperating group of biochemists was involved directly into the design phase. In the evaluation phase the biochemists were asked to load their datasets of interest and to evaluate them. To cover this, we added the following text to the paper:

“These scenarios were selected and conducted by the domain experts from our cooperating group of protein engineers. The group involved into the design of our proposed tool as well as the selection of interesting case studies and evaluation of the final visualization consisted of seven researchers - one professor (head of the protein engineering group), two post-docs, and four PhD students. They were asked to use our tool to explore their datasets and to evaluate the intuitiveness, understandability, benefits, and drawbacks of the proposed tool.”

1. *The simplification of the trajectories may not necessarily be the best choice. I would have opted for a multiresolution approach, where small details are successively omitted, thus effectively applying a low-pass filtering. Then, a simple slider could be used to intuitively and interactively choose the amount of simplification desired.*

The majority of discussions with the biochemists during the design phase was related to the selection of the best simplification method. There are plenty of possible approaches in the literature which could be adopted and could be potentially beneficial. We also tried to implement some of them and finally agreed on the approach described in the paper. However, as we also concluded that there is no ideal solution to this problem, we decided to combine the automatic approach with the interactive one so the user can decide on the level of simplification in different parts of the trajectory. The interactive simplification in fact supports the manual selection of the amount of simplification by using a slider.

1. *Another parameter is the window chosen for smoothing the scatterplots. Here also the question is how much is desired and whether the control should be given to the user.*

The sliding window parameter (functionality) was explicitly requested by domain experts.  
  
TODO: we should discuss among ourselfs the visibility of this parameter in GUI

|  |
| --- |
|  |

JP: To je napisane dost nejasne v texte. "To eliminate the noise in the data ..." chceme vyhladit data cez rychlu interakciu ... nieco na ten styl

TOTO prosím upravte, moc se nevyznám v tom, co tím chcete říct

1. *The analysis starts with an overview of the entire trajectory. This aspect does not scale to very large number of time steps. Any thoughts on this?*

Actually the overview representation was required and highly appreciated by the domain experts. It gives them the idea about the overall behavior of the ligand which is impossible to capture from the original data.

As for the scalability, we are able to compute the overview for simulation consisting of 50.000 time steps in order of minutes. The overview is actually the main reason why we are able to explore the long dynamic in reasonable time.

The overview breaks up the whole simulation into smaller parts which can be then evaluated in more details (individually) if and only if requested by domain expert.   
  
To compute the overview itself we do not need to store anything except information about resulting intervals in which ligand was moving towards or from the active site or whether was outside the molecule entirely (i.e., RAM is irrelevant). Since we evaluate each time step individually, I believe that the algorithm has linear scalability, i.e., if 50K takes 4min then 100K would take 8min etc. Moreover, this process could be easily parallelized and hence on my i7 (4 cores -> 8 threads) it would actually take just about 30s to solve it.  
  
The problems can occur when the ligand would move in a singe direction through the whole simulation and therefore the overview would consist of since interval (or user would mark the whole overview for detailed exploration). Then we need to evaluate whole simulation (computing nearby amino acids, tunnel radius, etc.).In this case various information has to be stored somewhere in order to visualise it. Nevertheless, it's again not that big issue. During my test the whole Analyst took about 1GB of RAM. The analysis itself is quite computation heavy task, it took 54:32 minutes to finish. Nevertheless, most attributes we evaluate can be computed individually (and the rest needs only a small neighbourhood). This means that we can again easily parallelize the algorithm.

Maybe it would be worth to mention also that we are buffering the simulation from the HDD, hence we can load infinitely long simulation without storing it in RAM

JP: Detaily ohladom konkretneho zrychlenia by som tam asi nedal, staci povedat, ze je to mozne urobit paralelne. Inac je to super.

AJ: Možný problém je určení, zda je ligand vně, na nebo uvnitř molekuly. Tady počítáme povrch (cca 2 s), ale pouze pro jeden snímek. Byl bych pro uvedení, že to děláme tak, že počítáme průměrnou strukturu a její povrch pak používáme.

1. *The Direction parameter sounds like a binary value. I assume it is more like the derivative of the distance, which would make a lot of sense, but it is not so clear from the description.*

The reviewer is right, we corrected the description of the Direction parameter in the following way:

“The direction parameter is a binary value computed as the derivative of the \textit{distance} attribute. In other words, we simply evaluate the ligand distance from the active site in two subsequent time steps and if the difference of the obtained values is positive we claim that ligand is moving towards active site and vice versa.”

1. *I wonder about the color choices in the overview visualization. Why were the colors chosen as they are. There is no obvious intuitive interpretation for me and green and blue are hard to distinguish. I actually don't see any blue on my print-out.*

At first, we intended to choose such colors which would be dissimilar to the color schemes used for encoding the physico-chemical properties of amino acids. Thus we have chosen pastel shaded colors. After further examination of multiple cases, this scheme is completely inapplicable. We would like to thank to the reviewer for this valuable comment.

In this second review, we have chosen a diverging and colorblind safe color scheme that is visually leading the user to the interesting areas (when is the ligand at the active site). This color scheme is slightly similar to the encoding of partial charge of amino acids. Nevertheless, we do not consider this to be a critical problem since the visualization supports the differentiation of the context in both cases.

1. *The description of the coloring of the line charts also remained unclear to me.*

We were not completely sure about what is meant by "line charts" in this comment thus we will adress the coloring of both bar charts and line representation of adjacent amino acids. The bar charts which are located straight under the overview visualization are colored according to categories of ligand movement. If particular parts of the trajectory are selected, then only these parts remain colored and the rest of the trajectory becomes gray to support the user to focus on interesting areas. The line representation showing ligand-lining amino acids is colored according to the selected property of adjacent amino acids. The color scheme for every property was given by the biochemists and is consistently used in the whole CAVER Analyst tool containing the implementation of our approach.

**Reviewer 2**

1. *This is well written paper and on a specific domain problem. I really don't have any major comments but I'd encourage authors to make their tool and code publicly available for reproducibility and impact.*

The tool will be publicly available via next version of CAVER Analyst (downloadable at www.caver.cz). Before its release, we will be happy to provide the alpha version of the tool on demand.

Minor:

1. *Why not use focus+context line graphs as opposed to scatterplots for visualizing attributes along paths. Using scatterplots seems a strange choice, particularly after trajectories are smoothed.*

We discussed this issue and we concluded that line graphs would be appropriate in cases when one of the depicted attributes is time (or any other attribute with linear progression). For instance, if we would try to visualize Hydrophobicty vs. Speed, we would most likely get a mess of lines going from one side of the graph to the other without any reasonable meaning. Therefore, we believe that the scatterplot representation is more robust in our case.

1. *The way the terms manual and automatic are used is a bit confusing, given the automatic simplification uses the manual simplification.  Maybe, consider using "interactive simplification" as opposed to "manual simplification" and refer Algorithm 1 directly while discussing automatic simplification*.

Adam: Tuhle změnu v článku bych zkusil udělat.

TODO udělat prosím

1. *The current formulation of trajectory complexity c(x) suggests that acute alpha's will lead to less complex trajectories, which isn't necessarily accurate. If c(x)is a curvature measure, then what you should care is the deviation of alpha from PI*.

Adam: Je to špatně popsané. Opravím to v článku a upravím tady komentář.

TODO opravit prosím

**Reviewer 4**

1. *While the proposed technique of simplification is interesting, it is not particularly novel and concerns with path simplification in such domains are not clearly addressed by the authors. The expert involvement is minimal and some more insights/comments would have been helpful.*

Adam: Já mám pouze domněnku, že jinak by byl velký problém z těch dat něco zjistit. To už jsme ale někam nejspíše do článku napsali.

*JP: (druhá věta)* Toto je vseobecny blabol. Kludne by som pouzil pouzil odvolavku na predchadzajuceho reviewera (resp. na nasu odpoved jemu).

*TODO komentovat to prosím*

1. *Some of the concerns that are not discussed in the paper are related to the simplification algorithm and the presentation of the simplified path. Based on our experience, domain experts are not particularly delighted about the use of simplification. The proposed manual simplification seems to simplify on top of the proposed automated simplification. There may be situations where an expert may want to see more fine grained details and constant simplification throughout may not be the answer.*

The reviewer is right that generally the domain experts are not fond of any simplification and prefer to explore the original data. However, in this case they admitted that some sort of simplification is necessary to get at least a rough idea about the ligand movement. The original data are too scattered to be able to understand the significant moments from it. Of course, the user can anytime return to the original data so the simplification can be taken only as a guidance through the data. Moreover, by enabling the interactive simplification on selected parts of the trajectory, the user can adjust the trajectory according to his or her needs.

1. *With regards to path/trajectory/line simplification, there is a classic algorithm called the Douglas-Peucker algorithm. I would recommend that the authors take a look at the algorithm and the following improvements on the algorithm.*

*Visvalingam, Mahes, and J. Duncan Whyatt. "The Douglas‐Peucker Algorithm for Line Simplification: Re‐evaluation through Visualization." Computer Graphics Forum. Vol. 9. No. 3. Blackwell Publishing Ltd, 1990.*

*Hershberger, John Edward, and Jack Snoeyink. Speeding up the Douglas-Peucker line-simplification algorithm. University of British Columbia, Department of Computer Science, 1992.*

TODO

1. *The authors may find the following paper relevant to the visualization in Figure 7 interesting:*

*Rodgers, Peter, Gem Stapleton, and Peter Chapman. "Visualizing sets with linear diagrams." ACM Transactions on Computer-Human Interaction (TOCHI) 22.6 (2015): 27.*

We thank the reviewer for this valuable comment. The suggested paper postulates six questions regarding the linear diagrams visualization (and answers them via a user study). We went through the paper and updated our visualization accordingly. Most of the techniques mentioned in the paper were already presented in our visualization except two of them. We have added the guidelines to indicate the start and end of overlaps in our visualization since the paper proved that study participants performed significantly better when using these guidelines. On the other hand, we cannot lower the amount of segments since in our case the time dependency is more important. Hence we cannot change the segments placement in order to minimize their number.

1. *Overall the paper presents an interesting application of a particular path simplification algorithm to a problem. The domain expert interaction is very minimal. For a paper such as this which is so heavily focused on an application domain, it is critical to have a significant section of the paper discussing domain expert evaluation of the proposed technique. The paper currently has an "Analysis Procedure and Discussion" section, but it is not clear how much interaction the domain expert had and whether any deeper insights were found. Usually, it also helps to show that the technique was applicable in more than one specific case.*

To address this issue, we added the information about the cooperating group of domain experts and how they participated on the design and evaluation. Moreover, we added another scenario of possible usage of our tool – the second scenario describes more complex case when three ligands are entering a protein structure.